

21.02.03

日 本 国 特 許 庁

JAPAN PATENT OFFICE

REC'D 24 APR 2003

WIPO

PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 5月16日

出 願 番 号

Application Number:

特願2002-141390

[ST.10/C]:

[JP2002-141390]

出 願 人

Applicant(s):

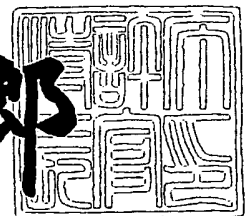
科学技術振興事業団  
株式会社国際電気通信基礎技術研究所

PRIORITY  
DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

2003年 4月 1日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3022816

BEST AVAILABLE COPY

【書類名】 特許願

【整理番号】 0020001

【特記事項】 特許法第30条第1項の規定の適用を受けようとする特  
許出願

【提出日】 平成14年 5月16日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 11/00

【発明者】

【住所又は居所】 埼玉県川口市本町4丁目1番8号 科学技術振興事業団  
内

【氏名】 パーハム モクタリ

【発明者】

【住所又は居所】 京都府相楽郡精華町光台二丁目2番地2 株式会社国際  
電気通信基礎技術研究所内

【氏名】 ニック キャンベル

【特許出願人】

【識別番号】 396020800

【氏名又は名称】 科学技術振興事業団

【特許出願人】

【識別番号】 393031586

【氏名又は名称】 株式会社国際電気通信基礎技術研究所

【代理人】

【識別番号】 100099933

【弁理士】

【氏名又は名称】 清水 敏

【手数料の表示】

【予納台帳番号】 173131

【納付金額】 21,000円

【提出物件の目録】

【物件名】	明細書	1
【物件名】	図面	1
【物件名】	要約書	1
【プルーフの要否】	要	

【書類名】 明細書

【発明の名称】 音声波形の特徴を高い信頼性で示す部分を決定するための装置およびプログラム、音声信号の特徴を高い信頼性で示す部分を決定するための装置およびプログラム、ならびに擬似音節核抽出装置およびプログラム

【特許請求の範囲】

【請求項 1】 音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置であって、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高信頼性部分として決定するための手段とを含む、音声波形の特徴を高い信頼性で示す部分を決定するための装置。

【請求項 2】 前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項 1 に記載の装置。

【請求項 3】 前記推定手段は、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第 1 の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第 2 の算出手段と、

前記第 1 の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第 2 の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、請求項 1 に記載の装置。

【請求項 4】 前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項 1 ～請求項 3 のいずれかに記載の装置。

【請求項 5】 音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置であって、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するための手段と、

前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む、擬似音節核抽出装置。

【請求項 6】 音声信号の特徴を高い信頼性で示す部分を決定するための装置であって、

前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて

、前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と

、  
第1の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第2の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、音声信号の特徴を高い信頼性で示す部分を決定するための装置。

【請求項7】 音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置としてコンピュータを動作させるプログラムであって、前記装置は、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高信頼性部分として決定するための手段とを含む、音声波形の特徴を高い信頼性で示す部分を決定するためのプログラム。

【請求項8】 前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極

小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項 7 に記載のプログラム。

【請求項 9】 前記推定手段は、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第 1 の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第 2 の算出手段と、

前記第 1 の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第 2 の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、請求項 7 に記載のプログラム。

【請求項 10】 前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項 7 ～請求項 9 のいずれかに記載のプログラム。

【請求項 11】 音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置としてコンピュータを動作させるプログラムであって、前記擬似音節核抽出装置は、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するための手段と、

前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記

有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む、擬似音節核抽出プログラム。

【請求項12】 音声信号の特徴を高い信頼性で示す部分を決定するための装置としてコンピュータを動作させるプログラムであって、前記装置は、

前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて、前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と

、  
第1の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第2の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、音声信号の特徴を高い信頼性で示す部分を決定するためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、一般的には音声波形からその特徴を高い信頼性で示す部分を抽出するための技術に関し、特に、音声波形の発生源の状態を高い信頼性で推定するために有効な領域を、音声波形から抽出するための技術に関する。

【0002】

【従来の技術】

【用語の定義1】

最初に、この節で使用される用語について定義する。

【0003】

「緊張音」(pressed sound)とは、発声の際に声門が緊張して



いるために声門を気流が通過しにくく、かつ通過をする際の気流の加速度が大きくなるように発声される音のことをいう。この場合、声門気流波形はサインカーブから大きく変形し、その微分波形の傾きが局部的に大きくなる。音声がこうした特徴を有する場合、「緊張性」の音声であると呼ぶことにする。

## 【0004】

「氣息音」(breathy sound)とは、発声の際に声門に緊張がないために気流が通過しやすく、その結果声門気流波形がサインカーブに近くなるように発声される音をいう。この場合、声門気流波形の微分波形の傾きが局部的に大きくなることはない。音声がこうした特徴を有する場合、「氣息性」の音声であると呼ぶことにする。

## 【0005】

「地声」(モーダル、modal)とは、緊張音と氣息音との中間の発声のことをいう。

## 【0006】

「AQ指数」(Amplitude Quotient)とは、声門(声帯)気流の波形のピークツーピークの振幅を、声門気流の波形の微分の振幅の最小値で除した値のことをいう。

## 【0007】

## [従来の技術]

音声認識と並んで重要な音声研究分野に、音声合成がある。最近の信号処理技術の発達により、音声合成が既に多くの分野で利用されている。しかし、これまでの音声合成は単にテキスト情報を音声化しているだけともいえ、人間が発話する際のような微妙な感情の表現までは行なえない。

## 【0008】

たとえば、人間が発話する際には、怒り、喜び、および悲しみなどの情報が、発話内容以外の情報、つまり声色などにより伝達される。このように発話に付随する、言語以外の情報をパラ言語情報と呼ぶ。これらはテキスト情報のみでは表わせない情報である。しかし従来の音声合成では、こうしたパラ言語情報を伝達することは難しかった。マンマシンインタフェースをより効率的なものとするた

めには、テキスト情報だけではなくパラ言語情報も音声合成の際に伝えられるようにすることが望ましい。

#### 【0009】

こうした問題を解決するために、種々の発話スタイルで連続的に音声合成を行なおうとする試みがある。ひとつの具体的な方策として次のようなものがある。すなわち、発話を録音してデータ処理可能な形でデータベース化し、さらにその中で所望の特徴(怒り、喜び、悲しみなど)を表わすと思われる発話単位にそれらの特徴を示すラベルを付ける。音声合成の際には所望のパラ言語情報に対応したラベルが付けられた音声を利用する。

#### 【0010】

しかし、十分な広さの発話スタイルをカバーできるようにデータベースを構築しようとするれば、膨大な量の録音音声を処理しなければならない。そのために、自動的にオペレータの介在なく確実にそうした特徴の抽出とラベル付け処理とを行なえるようにする必要がある。

#### 【0011】

以下、パラ言語情報の一例を挙げる。発話スタイルの一つとして、緊張音と氣息音という区別がある。緊張音では声門が緊張しているために、どちらかという強い発声となる。一方氣息音では、音声はサインカーブに近く、強いという印象はない。したがって緊張音と氣息音という区別も発話スタイルの一つとして重要であり、その程度を数量化できれば、パラ言語情報として利用できる可能性がある。

#### 【0012】

緊張音と氣息音との音質を区別する音響学的な指標については、今までにも数多くの研究がなされてきた。たとえば文末にリストした参考文献1を参照されたい。しかし、そうした研究の多くは、持続的に安定して母音を発音している間に録音された発話(または歌)を対象としたものに限定されていた。実際、膨大な量の発話の録音データから得られた音響測定データに基づいて、緊張性と氣息性との程度を信頼性高く計量しなければならないというのは非常に大きな問題であり、かつ実現された場合には非常に有用となるであろう。

## 【0013】

スペクトルドメインでの音源の属性を推定しようとする様々な手段が提案されて来たが、それよりも直接的な推定が、声門気流の波形とその導関数との組み合わせによって得られるはずである。そうした推定の一例が文末の参考文献2において提案されたAQ指数である。

## 【0014】

参考文献2では、AQ指数の一つの利点として、音圧レベル(SPL)から比較的独立していること、およびその値が主として発音の質的なものに依存していることがあげられている。他の利点として考えられるのは、このパラメータが純粹に振幅ドメインのものであって、種々の発話スタイルに応じた、推定された声門波形の時間ドメインの特徴量を測定する際の誤差源に対して比較的免疫性があることである。また、参考文献2の著者らによれば、様々な発音スタイルで「a」という母音を持続して発音した場合、4人の男性と4人の女性との全てに対して、発音を氣息性のものから緊張性に変えていくにしたがって、AQ指数の値は単調に減少したとのことである(参考文献2の第136頁)。したがってAQ指数は、ここで我々が解決しようとしている問題に関して有効である可能性が高い。ただし、AQ指数が有効となるためには、次の条件が満足される必要がある。

## 【0015】

- 1) 録音された通常の発話について、ロバストでかつ信頼性高くAQ指数を測定できること、および
- 2) そうした条件で測定された知覚上の特徴が顕著な部分を確認することができること。

## 【0016】

## 【発明が解決しようとする課題】

このような条件を満足させるためには、自然に発話された音声などの物理量を表わす音声波形から、いかにして信頼性高く音声波形の特徴を表わすパラメータを抽出できるかが重要である。特に音声の場合のように、発話が話者によりその細部まで完全にはコントロールされているわけでない場合、また様々な人が様々なスタイルで発話する場合には、パラメータを抽出すべき部分として信頼性がお

ける場所と、そうでない場所とが存在することが考えられる。そのため、音声波形のうちのどの部分を処理対象とするかが重要である。またそのために、日本語のように音節が発音の単位となる場合、音節の中心部（仮にこれを「音節核」と呼ぶ。）を誤りなく抽出できるようにすることが必要である。

## 【0017】

したがって、本発明の目的は、音声波形の特徴を高い信頼性で示す部分を決定することを可能とすることである。本発明のほかの目的は、本発明のさらに他の目的は、音節核を高い信頼性で抽出できるようにすることである。

## 【0018】

## 【課題を解決するための手段】

本発明の第1の局面は、複数の節に分解可能な、物理的量を表わす音声波形のデータに基づいて、音声波形の特徴を高い信頼性で示す部分を決定するための装置と、そうした装置としてコンピュータを動作させるプログラムに関する。この装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち、音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、データから音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、音声波形のうち、その変化が発生源により良好に制御されている領域を推定するための推定手段と、推定手段の出力と、発生源によって安定して発生されている領域として抽出手段により抽出され、かつ発生源によってその変化が良好に制御されていると推定手段によって推定された領域を音声波形の高信頼性部分として決定するための手段とを含む。

## 【0019】

抽出手段による抽出結果と、推定手段による推定結果との双方に基づいて音声波形の高信頼性部分を決定するので、決定結果がより確実なものとなる。

## 【0020】

抽出手段は、データに基づいて、音声波形の各区間が有声区間か否かを判定するための有声判定手段と、音声波形の所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で音声波形を節に分離するための手段と、音声波形のうち、

各節内で、当該節内のエネルギーのピークを含み、かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含んでもよい。

【0021】

有声と判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域が抽出されるので、発話者が安定して発声している区間を確実に抽出できる。

【0022】

また好ましくは、推定手段は、音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、データを用いて、線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第1の算出手段と、線形予測手段の出力に基づいて、音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第2の算出手段と、第1の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸上の分布と、第2の算出手段により算出された音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域を推定するための手段とを含む。

【0023】

フォルマント周波数の推定値の非信頼性と、音声波形の時間軸上のスペクトル変化の局所的な分散との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域が推定される。振動変化の発生源（たとえば発話者）が明確な意図をもって振動を制御している領域が推定できるので、そうした領域から振動の特徴量を算出すれば、算出された特徴量の信頼性が高くなることが期待できる。

【0024】

決定するための手段は、推定手段により音声波形の変化が発生源により良好に制御されていると推定された領域のうち、抽出手段により抽出された領域に含まれる領域を音声波形の高信頼性部分として決定するための手段を含んでもよい。

【0025】

音声波形の変化が発生源により良好に制御されていると推定された領域であって、かつ発生源により音声波形が安定に発生されているもののみを高信頼性部分として決定する。したがって真に信頼性が高い部分を抽出できる。

## 【0026】

本発明の他の局面は、音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置と、そうした装置としてコンピュータを動作させるプログラムとに関する。この擬似音節核抽出装置は、音声信号の各区間が有声区間か否かを判定するための有声判定手段と、音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で音声信号を擬似音節に分離するための手段と、音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む。

## 【0027】

有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域が擬似音節の核として抽出されるので、発話者が安定して発声しているときの音声抽出することができる。

## 【0028】

本発明のさらに他の局面は、音声信号の特徴を高い信頼性で示す部分を決定するための装置と、そうした装置としてコンピュータを動作させるプログラムとに関する。当該装置は、音声信号に対する線形予測分析を行なうための線形予測手段と、線形予測手段によるフォルマントの推定値と、音声信号とに基づいて、フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第1の算出手段と、線形予測手段による線形予測分析の結果に基づいて、音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第2の算出手段と、第1の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸上の分布と、第2の算出手段により算出された音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域を推定するための手段とを含む。

## 【0029】

フォルマントの推定値の非信頼性の時間軸上の分布も、音声信号のスペクトル変化の局所的な分散の時間軸上の分布も、その極小部ではいずれも音声信号のうちでその発生源により音声波形の発生が良好に制御されている部分を示す。これらの双方を用いて領域を推定するので、音声波形の発生が良好に制御されている部分を信頼性高く特定することができる。

## 【0030】

## 【発明の実施の形態】

以下に述べる本発明の実施の形態は、コンピュータおよびコンピュータ上で動作するソフトウェアにより実現される。もちろん、以下に述べる機能の一部又は全部を、ソフトウェアでなくハードウェアで実現することも可能である。

## 【0031】

## 〔用語の定義2〕

以下、本実施の形態の説明で使用される用語について定義する。

## 【0032】

「擬似音節」とは、音声信号から所定の信号処理によって決定される信号の切れ目のことを指し、日本語音声の場合の音節を推定したものに対応する。

## 【0033】

「ソノラントエネルギー」とは、音声信号のうちで、所定周波数（たとえば60Hz～3kHzの周波数領域）のエネルギーのことをいい、デシベルで表わされる。

## 【0034】

「信頼性の中心」(center of reliability)とは、音声波形に対する信号処理の結果、音声波形のうちで、対象となる音声波形の特徴を信頼性高く抽出できるとみなされることとなった領域のことをいう。

## 【0035】

「ディップ」とは、グラフなどの図形がくびれた部分のことをいう。特に、時間の関数として変化するような値の時間軸上の分布により形成される波形のうち

、極小値に対応する部分をいう。

【0036】

「非信頼性」とは、信頼性のなさを表わす尺度のことをいう。非信頼性は信頼性の逆の概念である。

【0037】

図1に、本実施の形態で利用されるコンピュータシステム20の外観図を、図2にコンピュータシステム20のブロック図を、それぞれ示す。なおここに示すコンピュータシステム20はあくまで一例であり、この他にも種々の構成が可能である。

【0038】

図1を参照して、コンピュータシステム20は、コンピュータ40と、いずれもこのコンピュータ40に接続されたモニタ42、キーボード46、およびマウス48を含む。コンピュータ40にはさらに、CD-ROM (Compact Disc Read-Only Memory) ドライブ50と、FD (Flexible Disk) ドライブ52とが内蔵されている。

【0039】

図2を参照して、コンピュータシステム20はさらに、コンピュータ40に接続されるプリンタ44を含むが、これは図1には示していない。またコンピュータ40はさらに、CD-ROMドライブ50およびFDドライブ52に接続されたバス66と、いずれもバス66に接続された中央演算装置 (Central Processing Unit: CPU) 56、コンピュータ40のブートアッププログラムなどを記憶したROM (Read-Only Memory) 58、CPU56が使用する作業エリアおよびCPU56により実行されるプログラムの格納エリアを提供するRAM (Random Access Memory) 60、および後述する音声データベースを格納したハードディスク54を含む。

【0040】

以下に述べる実施の形態のシステムを実現するソフトウェアは、たとえば、CD-ROM62のような記録媒体上に記録されて流通し、CD-ROMドライブ



50のような読取装置を介してコンピュータ40に読込まれ、ハードディスク54に格納される。CPU56がこのプログラムを実行する際には、ハードディスク54からこのプログラムを読み出してRAM60に格納し、図示しないプログラムカウンタによって指定されるアドレスから命令を読み出して実行する。CPU56は、処理対象のデータをハードディスク54から読出し、処理結果を同じくハードディスク54に格納する。

#### 【0041】

コンピュータシステム20の動作自体は周知であるので、ここではその詳細については繰り返さない。

#### 【0042】

なお、ソフトウェアの流通形態は上記したように記憶媒体に固定された形には限定されない。たとえば、ネットワークを通じて接続された他のコンピュータからデータを受取る形で流通することもあり得る。また、ソフトウェアの一部が予めハードディスク54中に格納されており、ソフトウェアの残りの部分をネットワーク経由でハードディスク54に取込んで実行時に統合するような形の流通形態もあり得る。

#### 【0043】

一般的に、現代のプログラムはコンピュータのオペレーティングシステム(OS)によって提供される汎用の機能を利用し、それらを所望の目的にしたがって組織化した形態で実行することにより前記した所望の目的を達成する。したがって、以下に述べる本実施の形態の各機能のうち、OSまたはサードパーティが提供する汎用的な機能を含まず、それら汎用的な機能の実行順序の組合せだけを指定するプログラム(群)であっても、それらを利用して全体的として所望の目的を達成する制御構造を有するプログラム(群)である限り、それらが本発明の技術的範囲に含まれることは明らかである。

#### 【0044】

本実施の形態のプログラムを装置とみなして機能的に示したのが図3以下のブロック図である。図3を参照して、この装置80は、ハードディスク54に格納された音声データ82に対して以下に説明する処理を行なって、音声データに含

まれる各処理単位（たとえば音節）ごとに前述したA Q指数を算出し出力するためのものである。なお、音声データは後述するように1フレーム32 msecとなるように予めフレーム化されている。

#### 【0045】

装置80は、音声データに対して高速フーリエ変換（Fast Fourier Transform: FFT）を行なうFFT処理部90と、FFT処理部90の出力を用い、音声データにより表わされる音声波形のうちの60 Hz～3 kHzの周波数領域のエネルギーの時間的変化および音声のピッチの変化に基づいて、音声データにより表わされる音声波形の各音節のうち、話者の発声機構によって安定して発生されている領域（これを以後「擬似音節核」と呼ぶ。）とを抽出する音響・韻律分析部92と、音声データ82に対してケプストラム分析を行ない、さらに、FFT処理部90の出力を用いてケプストラム分析の結果音声スペクトルの変化が少なく、音声データの特徴を信頼性高く抽出できると思われる部分（これを「高信頼性・小変動部の中心」または「高信頼・小変動の中心」または単に「信頼性の中心」と呼ぶ。）を推定するためのケプストラム分析部94とを含む。

#### 【0046】

装置80はさらに、ケプストラム分析部94の出力する信頼性の中心（高信頼性・小変動部の中心）の中で、音響・韻律分析部92の出力する擬似音節核の中にあるものだけを擬似音節中心として抽出するための擬似音節中心の抽出部96と、擬似音節中心の抽出部96によって抽出された擬似音節中心に対応する音声データに対して、フォルマントの初期推定と最適化処理とを行なって最終的なフォルマントの推定値を出力するためのフォルマントの最適化部98と、音声データに対して、フォルマントの最適化部98から出力されるフォルマント値を用いた適応的フィルタ処理などの信号処理を行なって声門気流波形の微分を推定し、さらにそれを積分することによって声門気流波形を推定し、それらに基づいてA Q指数を計算するためのA Q指数計算部100とを含む。

#### 【0047】

図4は、音声データの構成を模式的に示す図である。図4を参照して、音声デ

一タ波形102は、それぞれ32 msecごとのフレームに分けられ、かつ前後のフレーム間では8 msecごとにずらしてデジタル化されている。そして、後述する処理では、たとえばある時点 $t_0$ では第1のフレームを先頭として処理をし、次の時点 $t_1$ では8 msecずれた次の第2のフレームを先頭として処理をする、という形で処理を行なっていく。

## 【0048】

図5は、図3に示す音響・韻律分析部92のブロック図である。図5を参照して、音響・韻律分析部92は、音声波形から測定される音源のピッチを用いて、処理対象のフレームが有声区間か否かを判定する（この方法については参考文献3を参照）ためのピッチ判定部110と、FFT処理部90の出力に基づいて所定周波数領域（60 Hz～3 kHz）のソノラントエネルギーの時間軸上の波形分布を算出するためのソノラントエネルギー算出部112と、ソノラントエネルギー算出部112によって算出されるソノラントエネルギーの時間軸上の分布波形の輪郭に対して凸包アルゴリズムを適用することにより、ソノラントエネルギーの時間軸上の分布波形の輪郭の中のディップを検出して、入力音声を擬似音節に分割する（この方法については参考文献4および5を参照）ためのディップ検出部114と、ディップ検出部114によって得られた擬似音節中の、ソノラントエネルギーの最大値（SE peak）が得られる点を起点として、その左右に、ソノラントエネルギーが所定のしきい値（ $0.8 \times \text{SE peak}$ ）より大きく、かつピッチ判定部110によって有声区間であると判定されたフレームであって、かつ同じ擬似音節中のフレームを1フレームずつ広げていくことにより、擬似音節核を出力するための有声・エネルギー判定部116とを含む。

## 【0049】

図6は、図3に示すケプストラム分析部94のブロック図である。図6を参照して、ケプストラム分析部94は、音声データ82の音声波形に対して選択的線形予測（Selective Linear Prediction: SLP）分析を行なって、SLPケプストラム係数 $c_{f,i}$ を出力するための線形予測分析部130と、このケプストラム係数に基づいて先頭の4つのフォルマンツの周波数と帯域との初期推定値を算出するためのフォルマンツ推定部132とを含む。

フォルマント推定部 132 は、参考文献 6 により提案された線形ケプストラムフォルマントマッピングを利用し、かつ同じデータのサブセットを使用して注意深く測定された母音フォルマントに対するマッピングを学習させてある。この学習については、参考文献 7 を参照されたい。

## 【0050】

ケプストラム分析部 94 はさらに、推定されたフォルマント周波数などに基づいてケプストラム係数  $C_i^{\text{simp}}$  を再計算するためのケプストラム再生部 136 と、FFT 処理部 90 の出力に対して対数変換およびコサイン逆変換 (IDCT) を行なって FFT ケプストラム係数を算出するための対数変換および逆 DCT 部 140 と、ケプストラム再生部 136 により計算されたケプストラム係数  $C_i^{\text{simp}}$  と、対数変換および逆 DCT 部 140 により計算された FFT ケプストラム係数  $C_i^{\text{FFT}}$  との間の差を表わす値として次の式により定義されるケプストラム距離  $d_f^2$  を計算し、フォルマント推定部 132 によって推定されたフォルマント周波数などの値の非信頼性を表わす指標として出力するためのケプストラム距離計算部 142 とをさらに含む。

## 【0051】

【数 1】

$$d_f^2 = \text{Sum}_i \{ i^2 \cdot (c_i^{\text{simp}} - c_i^{\text{FFT}})^2 \}$$

フォルマント推定部 132、ケプストラム再生部 136、ケプストラム距離計算部 142、および対数変換および逆 DCT 部 140 により、線形予測分析の結果に基づいて推定されたフォルマント周波数などの値の非信頼性が算出される。

## 【0052】

ケプストラム分析部 94 はさらに、線形予測分析部 130 の出力するケプストラム係数から  $\Delta$  ケプストラムを算出する為の  $\Delta$  ケプストラム算出部 134 と、 $\Delta$  ケプストラム算出部 134 の出力する  $\Delta$  ケプストラムに基づいて、各フレームごとに、そのフレームを含む 5 フレームのスペクトル変化の大きさの分散を算出する為のフレーム間分散算出部 138 とを含む。フレーム間分散算出部 138 の出

力は、局所的なスペクトルの動きの時間軸上の分布波形の輪郭を表わすものとなり、その極小値は、参考文献8で提案されている調音音声学理論にならっていえば、制御された動きCM (Controlled Movement) を示すものと考えることができ。

## 【0053】

さらにケプストラム分析部94は、ケプストラム距離計算部142の出力するフォルマント周波数の推定値の非信頼性を示す値と、フレーム間分散算出部138の出力する各フレームごとの局所的なフレーム間分散値とを受け、両者の値を規格化し統合して、フレームごとの音声信号の非信頼性を示す値の時間軸上の分布波形として出力するための規格化および統合部144と、規格化および統合部144の出力する非信頼性の値の時間軸上の分布波形により形成される波形の輪郭のディップを凸包アルゴリズムにより検出して、信頼性の中心候補として出力するための信頼性の中心候補出力部146とを含む。

## 【0054】

図7は、図6に示す規格化および統合部144のブロック図である。図7を参照して、規格化および統合部144は、ケプストラム距離計算部142により出力されたケプストラム距離を $[0, 1]$ の値に規格化するための第1の規格化部160と、フレーム間分散算出部138が各フレームごとに算出するフレーム間分散の値を $[0, 1]$ の値に規格化するための第2の規格化部162と、局所的なフレーム間分散の値の時間軸上の位置を、ケプストラム距離計算部142の出力するケプストラム距離のサンプリングタイミングと一致させるように線形補間処理を行なうための補間処理部164と、第1の規格化部160の出力と補間処理部164の出力とを1フレームごとに平均して出力するための平均計算部166とを含む。平均計算部166の出力は、統合された値の時間軸上の分布波形の輪郭を表わす。信頼性の中心候補出力部146によってこの波形の輪郭のディップ(極小部)を検出することにより、非信頼性が最も低い部分(信頼性が最も高い部分)を信頼性の中心の候補として特定することができる。

## 【0055】

図8は、図3に示すフォルマントの最適化部98のブロック図である。図8を

参照して、フォルマントの最適化部 98 は、音声波形に対して F F T 処理を行なうための F F T 処理部 180 と、F F T 処理部 180 の出力に対して対数変換およびコサイン逆変換を行なうための対数変換および逆 D C T 部 182 と、対数変換および逆 D C T 部 182 の出力する F F T ケプストラム係数と、後述するフォルマントの推定値との間の距離を計算するためのケプストラム距離計算部 184 と、信頼性の中心候補の各々における第 1 ～第 4 のフォルマント周波数の初期推定値を初期値とし、ケプストラム距離計算部 184 が計算する距離を最小にするように山登り法によってフォルマントの推定値を最適化するための距離最小化処理部 186 とを含む。距離最小化処理部 186 によって最適化されたフォルマント推定値がフォルマントの最適化部 98 の出力として A Q 指数計算部 100 に与えられる。

#### 【0056】

図 9 を参照して、A Q 指数計算部 100 は、音声信号のうちで音節中心に相当する位置の 64 m s e c の部分のうち、70 H z 以上の周波数成分のみを選択的に通過させるためのハイパスフィルタ 200 と、ハイパスフィルタ 200 の出力のうち、最適化された第 4 フォルマント周波数とその帯域との和以下の周波数成分のみを選択的に通過させるための適応的ローパスフィルタ 202 と、適応的ローパスフィルタ 202 の出力に対し、第 1 ～第 4 フォルマント周波数を用いた適応的逆フィルタ処理を行なうための適応的逆フィルタ 204 とを含む。適応的逆フィルタ 204 の出力は、声門気流波形の微分波形となる。

#### 【0057】

A Q 指数計算部 100 はさらに、適応的逆フィルタ 204 の出力を積分して声門気流波形を出力するための積分回路 206 と、積分回路 206 の出力のピークツーピークの最大振幅を検出するための最大ピーク間振幅検出回路 208 と、適応的逆フィルタ 204 の出力の負のピークの最大振幅を検出するための最大の負のピーク振幅検出回路 210 と、最大の負のピーク振幅検出回路 210 の出力に対する最大ピーク間振幅検出回路 208 の出力の比を算出するための比計算回路 212 とを含む。比計算回路 212 の出力が A Q 指数である。

#### 【0058】

図1～図9に示した装置は以下のように動作する。まず、使用された音声データ82について説明する。この音声データは参考文献9で使用されたものであり、日本語のネイティブスピーカーである女性が3つの物語を読んだものを録音して作成されたものである。この物語は、怒りと、喜びと、悲しみという感情を引き起こすように予め作成されていたものである。物語の各々は400文の長さ（おおよそ30,000音素）以上の発話を含む。各発話は別々の音声波形ファイルに格納され処理された。

## 【0059】

各文の発話データはFFT処理部90によるFFT処理の後、以下のようにして処理される。処理は大きく見て二つの系統に分かれ実行される。第1の系統は音響・韻律分析部92で行なわれる音響韻律的な処理であり、他の系統はケプストラム分析部94が行なう音響音声学的な処理である。

## 【0060】

音響韻律的な系統の処理では、図5に示すソノラントエネルギー算出部112によって60Hz～3kHz周波数領域のソノラントエネルギーが算出される。ソノラントエネルギー算出部112の出力する一文の発話データの全体波形の輪郭から、ディップ検出部114が凸包アルゴリズムによりディップを検出する。このディップにより、この発話文は擬似音節に分割される。

## 【0061】

有声・エネルギー判定部116は、擬似音節の中でソノラントエネルギーが最大（SE peak）となる点を見つける。この点が擬似音節核の初期点である。有声・エネルギー判定部116はさらに、この擬似音節核の初期点から始めて、その左右に向かい、ソノラントエネルギーが $0.8 \times \text{SE peak}$ 以下のフレーム、またはピッチ判定部110が有声でないと判定したフレーム、または擬似音節の外のフレームに出会うまで、擬似音節核の範囲を広げる。こうして擬似音節核の境界が決定される。この情報は擬似音節中心の抽出部96に与えられる。なお、ここでしきい値として0.8の値を用いているが、これは単なる例であって、応用によりこのしきい値を適切な値に代える必要がある。

## 【0062】

図6を参照して、入力された一つの発話文に対して線形予測分析部130が線形予測分析を行ない、SLPケプストラム係数を出力する。Δケプストラム算出部134がこのSLPケプストラム係数に基づいてΔケプストラムを算出し、フレーム間分散算出部138に与える。フレーム間分散算出部138は、このΔケプストラム係数に基づき、各フレームごとに、そのフレームを含む5フレームの中での局所的なスペクトル変化の分散を計算する。この分散が小さいほど発話者の発声が発話者によりよく制御されていると考えられ、逆にこの分散が大きいと話者による制御がよくされていないと考えられるので、フレーム間分散算出部138の出力は発話者の発声が信頼されない程度（非信頼性）を表わすと考えられる。

## 【0063】

図6をさらに参照して、フォルマント推定部132は、線形ケプストラムフォルマントマッピングを用い、SLPケプストラム係数に基づいて第1～第4フォルマントの周波数と帯域とを推定する。ケプストラム再生成部136は、フォルマント推定部132により推定された第1～第4フォルマントに基づいて逆にケプストラム係数を算出しケプストラム距離計算部142に与える。対数変換および逆DCT部140は、フォルマント推定部132およびケプストラム再生成部136が処理したのと同じフレームのものの音声データに対して対数変換およびコサイン逆変換を行なってFFTケプストラム係数を算出しケプストラム距離計算部142に与える。ケプストラム距離計算部142は、ケプストラム再生成部136からのケプストラム係数と対数変換および逆DCT部140からのケプストラム係数との間の距離を前述の「数1」の式にしたがって計算する。この結果得られるのは、フォルマント推定部132が推定したフォルマントの非信頼性を示す値の時間軸上の分布を表わす波形と考えられる。ケプストラム距離計算部142は、この結果を規格化および統合部144に与える。

## 【0064】

図7を参照して、規格化および統合部144の第1の規格化部160は、図6のケプストラム距離計算部142の出力する、フォルマントの推定値から算出された各フレームごとの非信頼性値を[0, 1]の範囲に正規化して平均計算部1



66に与える。第2の規格化部162は、図6のフレーム間分散算出部138が出力する、フレームごとに計算された局所的なフレーム間分散の値を $[0, 1]$ の範囲に正規化して補間処理部164に与える。補間処理部164は、第2の規格化部162の各値に対し、第1の規格化部160の出力する各フレームのサンプリングポイントに対応する値が得られるように線形補間処理を行なって平均計算部166に与える。平均計算部166は、フレームごとに、第1の規格化部160の出力と補間処理部164の出力とを正規化し、その結果を時間軸上の非信頼性の分布を示す統合された波形として信頼性の中心候補出力部146に出力する。

#### 【0065】

信頼性の中心候補出力部146は、凸包アルゴリズムにより、規格化および統合部144の出力する統合された波形の輪郭のディップを検出して、そのフレームを特定する情報を図3の擬似音節中心の抽出部96に対して信頼性の中心の候補として出力する。

#### 【0066】

図3に示す擬似音節中心の抽出部96は、図6に示す信頼性の中心候補出力部146から与えられた信頼性の中心の中で、音響・韻律分析部92から与えられた擬似音節核の中にあるもののみを擬似音節中心として抽出する。

#### 【0067】

以上の処理によって、音声データのうちで音声データの特徴を抽出する、または音声データをラベル付けするために適した高信頼性・小変動領域を示す情報が得られたことになる。したがって、この情報によって特定されるフレームについて所望の処理を行なえばよい。本実施の形態の装置では、擬似音節中心の抽出部96はこの情報をフォルマントの最適化部98に与え、フォルマントの最適化部98はこの情報を用いて、以下のようにして擬似音節中心におけるAQ指数を算出する。

#### 【0068】

なお、本実施の形態の装置では、擬似音節中心の長さは連続する5フレームとする。1フレームは32 msecであり、連続するフレームは互いに8 msec

ずつつづれているから、5フレームの全体では64 msecの音声期間に相当する。

#### 【0069】

これらの擬似音節中心におけるAQ指数は、図9のAQ指数計算部100中で得られる声門気流の波形により直接計算することができる。しかし、声門気流の推定自体、もともとのフォルマントに相当する声道の共振によって影響されており、その信頼性は共振の影響をもとの音声波形の64 msecのデータから取り除くことができるかに依存している。したがって、そのような計算によって得られたAQ指数は信頼できないものとなる。

#### 【0070】

一方、擬似音節中心におけるフォルマントは、スペクトルがよく一致しているという意味で、既により推定となっているが、本実施の形態の装置では、さらに以下のようにしてフォルマント周波数を最適化する。

#### 【0071】

すなわち、図8を参照して、FFT処理部180は音声波形に対してフレームごとにFFT処理を行なう。対数変換および逆DCT部182はFFT処理部180の出力に対して対数変換およびコサイン逆変換を行なう。ケプストラム距離計算部184は、対数変換および逆DCT部182の出力するケプストラム係数と距離最小化処理部186から与えられるケプストラム係数の推定値との間の距離を計算する。距離最小化処理部186は、フォルマントの推定値を表わすケプストラム係数の値を起点として、ケプストラム距離計算部184により計算される距離が最小値となるように山登り法によって距離最小化処理部186から与えられたケプストラム係数の値をさらに最適化し、最小値が得られるときのフォルマント推定値を出力する。

#### 【0072】

AQ指数計算部100の内部構成は図9に示されており、この図9を参照して、擬似音節中心における音声データはまずハイパスフィルタ200を通り、その結果70 Hz以下の低周波数の雑音が除去される。さらに適応的ローパスフィルタ202によって第4フォルマントより高い周波数領域のスペクトル情報が除去

される。そして、適応的逆フィルタ 2 0 4 によって第 1 ～第 4 フォルマントによる影響が除去される。

#### 【 0 0 7 3 】

その結果、適応的逆フィルタ 2 0 4 の出力は声門気流の波形の微分のよい推定値となる。これを積分回路 2 0 6 で積分することにより声門気流の波形の推定値が得られる。最大ピーク間振幅検出回路 2 0 8 によって声門気流の波形のピークツーピークの振幅の最大値を検出する。最大の負のピーク振幅検出回路 2 1 0 によって声門気流の微分波形のサイクル内での負の最大の振幅を検出する。最大ピーク間振幅検出回路 2 0 8 の出力の、最大の負のピーク振幅検出回路 2 1 0 の出力に対する比を比計算回路 2 1 2 で計算することにより、擬似音節中心における A Q 指数が得られる。

#### 【 0 0 7 4 】

こうして得られた A Q 指数は、各擬似音節中心におけるもとの音声データの特徴（緊張音－氣息音の間の度合い）を信頼性高くあらわしている。これら各擬似音節中心に対して A Q 指数を計算し、さらにこれら得られた A Q 指数を補間することにより、擬似音節中心以外の部分の A Q 指数を推定することもできる。そうすることにより、音声データのうち、一定の A Q 指数を示す部分に、当該 A Q 指数に対応した適切なラベルをパラ言語情報として付けておき、音声合成の際には、所望の A Q 指数を有する音声データを使用すれば、単なるテキストだけでなく、パラ言語情報をも含んだ形での音声合成を行なうことが可能になる。

#### 【 0 0 7 5 】

図 1 0 ～図 1 2 に、本実施の形態の装置をコンピュータにより実現した際の画面表示例を示す。

#### 【 0 0 7 6 】

図 1 0 を参照して、このプログラムによる表示ウィンドウには、音声データ波形 2 4 0 と、音声データに対して付された音声ラベル 2 4 2 と、基本周波数の波形の時間軸上の分布波形の輪郭 2 4 4 と、ソノラントエネルギーの変動の時間軸上の分布波形の輪郭 2 4 6 と、 $\Delta$ ケプストラムから計算されたスペクトル変化の局所的な分散の時間軸上の分布波形の輪郭 2 4 8 と、フォルマント－FFT ケプス

トラム距離の時間軸上の分布波形の輪郭 250 と、スペクトル変化の局所的な分散の分布波形の輪郭 248 およびフォルマントークプストラム距離の分布波形の輪郭 250 を統合した波形である非信頼性の時間軸上の分布波形の輪郭 252 と、上述のようにして算出された擬似音節中心での声門の A Q 指数 254 と、各擬似音節中心で推定された声道の面積関数 256 とが示されている。

## 【0077】

音声データ波形 240 の表示領域に示された太い縦線 232 と、ソノラントエネルギーの変動の輪郭 246 の表示領域に示された太い縦線とは擬似音節の境界を示す。音声データ波形 240 の表示領域に示された細い縦線 230 と、ソノラントエネルギーの変動の輪郭 246 および基本周波数の波形輪郭 244 の表示領域に示された細い縦線は擬似音節核の境界を示す。

## 【0078】

非信頼性の波形 252 の表示領域に示された縦線は波形の極小値部分（ディップ）であり、そこを中心として A Q 指数が計算されている部分が最も信頼性の高い部分である。なお A Q 指数が計算された期間および値は横棒で示されており、横棒の縦位置が高いほど緊張音に近く、低いほど氣息音に近い。

## 【0079】

図 11 には、図 10 の点線のボックス 262 で示される時点での声門気流波形の推定値 270 と、その微分波形 272 と、推定された声門気流波形のスペクトル 274 とが示されている。図 10 のボックス 262 に対応する時点では A Q 指数 254 は高く、すなわちこの時点の発声は緊張音に近い。図 11 に示すとおり、このときの声門気流の波形はのこぎり形に近く、サインウェーブの波形からは遠く異なっている。また、微分波形はすどく変化している。

## 【0080】

図 12 には、図 10 の点線のボックス 260 で示される時点での声門気流波形の推定値 280 と、その微分波形 282 と、推定された声門気流波形のスペクトル 284 とが示されている。図 10 のボックス 260 に対応する時点では A Q 指数 254 は低く、すなわちこの時点の発声は氣息音に近い。図 12 に示すとおり、このときの声門気流の波形はきれいなサインカーブに近い。微分波形も緩やか

なものとなっている。

#### 【0081】

上に述べた装置を用い、前述した音声データを実際に処理して擬似音節中心を抽出し、各擬似音節中心に対してA Q指数を算出しする一方、それらの擬似音節中心に対応する音を人間が聞いたときに感ずる感想と、A Q指数との相関を以下のようにして調査した。

#### 【0082】

上記した装置を用いて抽出された信頼性の中心は22,000個であり、その各々について対応する声門気流波形およびA Q指数と、もとの音声波形のRMS (Root Mean Square) エネルギー(dB)とを算出した。これら信頼性の中心のうち、同一の音節核中に存在しかつ互いのA Q指数がほぼ一致しているものをまとめ、さらにそれら信頼性の中心のうち、統合された非信頼性の値が0.2以上のものを棄却することにより、聴覚刺激として使用可能と思われる音節核の数は15,000をわずかに超えたものとなった。

#### 【0083】

このデータセットに対して算出された統計情報に基づき、知覚上の評価を行なうために60の刺激からなるサブセットを選択した。具体的には、前述した3つの感情を表わすデータベースの各々について、極めて低い、または極めて高い、または各感情に対するA Q指数の平均値マイナスその分布の標準偏差( $\sigma$ )近辺、またはA Q指数の平均値プラス標準偏差近辺、の4つのカテゴリのいずれかにA Q指数が属するような信頼性の中心を含む音節核を5つずつ選択した。

#### 【0084】

このようにして選択された60個の擬似音節核の時間的長さは32 msecから560 msecの範囲であり、その平均は171 msecであった。通常の聴覚的能力を有する11人の被験者が、これら短時間の刺激の各々について聴覚的評価を行なった。被験者は静粛なオフィス環境で、高音質のヘッドフォンを用い、各刺激を必要な回数だけ聞き、各刺激について、それぞれ「気息性」および「強さ」とだけ説明した二つのスケールにしたがい、7段階で採点した。各被験者の採点は各々比例により[0, 1]の範囲に正規化され、正規化した点数に基づ

いて、60個の刺激の各々についての11人の被験者全ての気息性および強さに関する平均値を算出した。

#### 【0085】

図13は、上のようにして調べた気息性と、音響的に測定したAQ指数の値とを比較する散布図である。これら60対の値に対する線形相関係数は0.77であった。この相関は必ずしも高いものではないが、刺激に対するAQの測定値が高くなれば、その刺激に対して感じられる気息性も平均すれば高くなるという明らかな傾向があることを裏付けるものといえる。図13の散布図上で想定されるベストフィットの直線から最も遠い位置に存在する点のいくつかをより詳細に調べると、誤差の原因として次のようなものが浮かび上がる。すなわち、動的制約が欠如しているために生ずる、5つのフレーム中でのフォルマントの非連続性、5つのフレームに含まれていない音節核の一部において生ずる高い気息性、および5つのフレーム中の母音部分に対して、隣接した鼻音がおよぼす強い影響などである。

#### 【0086】

さらに、図13からは、中位から下位のAQ指数を有する刺激に対しては、気息性の感じ方が広いことに気づく。これは、気息性が低い刺激に気息性に関する点数をつけることが難しく、むしろ地声または緊張音的な発音という側面から点数付けしたほうがよりよく特徴を表わせるのではないか、という直感的な理解を裏付けるものと思われる。

#### 【0087】

ここでは図としては示していないが、強さの感じ方を、同じ信頼性の中心において測定されたRMSエネルギーと比較するための散布図も作成した。その相関係数は0.83となり、より高度な重み付けを用いて強さの感じ方を測定しているわけではないにもかかわらず、その関係の強さを裏付けるものとなっている。

#### 【0088】

以上のように本実施の形態では、音響・韻律的分析と、ケプストラム分析とを組合せて、(i)録音された自然な発声中の擬似音節の信頼性の中心の位置を決定するための、(ii)参考文献2で提案されたAQ指数により定量化された音

源の属性を測定するための、全くオペレータが介在する必要がない方法および装置を実現した。そして、その方法および装置を用いて行なった音声知覚の実験の結果は、擬似音節核中で知覚された気息性と強い相関を持つ、頑健性をもって測定できる値としてのAQ指数の重要性を確認するものであった。実際、前述したような誤差源が存在しているにもかかわらず、AQ指数と気息性の知覚との間に見出された相関により、音質パラメータとしてのAQ指数をさらに研究する必要があることを確認することができた。

【0089】

そしてこの方法および装置により、発声単位に対するパラ言語的なラベル付けを行なうことができる可能性が高くなる。そうした発声単位を用い、所望のラベル付けがされた発声単位を用いて音声の連続合成を行なうことにより、緊張音から地声、さらに気息的な発音までの範囲にわたる幅広い発声スタイルを用いたマンマシンインタフェースを実現することが可能となる。

【0090】

[参考文献]

- (1) Sundberg,  
J. (1987). The science of the singing voice, Northern Illinois University  
Press, Delcalb, Illinois.
- (2) Alku,  
P. & Vilkman, E. (1996). "Amplitude domain quotient for characterization of  
the glottal volume velocity waveform estimated by inverse filtering", Speech  
Comm., 18(2), 131-138.
- (3) Hermes,  
D. (1988). "Measurement of pitch by subharmonic summation", J. Acoust.  
Soc. Am.  
83(1), 257-264.

- (4) Mermelstein,  
P. (1975). "Automatic segmentation of speech into syllabic units", J.  
Acoust.  
Soc. Am. 58(4), 880-883.
- (5) Lea,  
W.A. (1980). "Prosodic aids to speech recognition", in Lea, W.A. (ed.)  
, Trends  
in Speech Recognition, Prentice-Hall, New Jersey, 166-205.
- (6) Broad,  
D.J. & Clermont, F. (1989). "Formant estimation by linear transformatio  
n of  
the LPC cepstrum", J. Acoust. Soc. Am. 86 (5), 2013-2017.
- (7) Mokhtari,  
P., Iida, A. & Campbell, N. (2001). "Some articulatory correlates of  
emotion variability in speech : a  
preliminary study on spoken Japanese vowels", Proc. Int. Conf. on Speec  
h  
Process., Taejon, Korea, 431-436.
- (8) Peterson,  
G.E., & Shoup, J.E. (1966). "A  
physiological theory of phonetics", J. Speech Hear. Res. 9, 5-67.
- (9) Iida,  
A., Campbell, N., Iga, S., Higuchi, F. & Yasumura, M. (1998). "Acoustic  
nature and perceptual testing of corpora of emotional speech", Proc. 5t  
h Int.  
Conf. on Spoken Lang. Process., 1559-1562.

【図面の簡単な説明】

【図 1】 本発明の一実施の形態のプログラムを実行するコンピュータシス  
テムの外観を示す図である。



【図 2】 図 1 に示すコンピュータシステムのブロック図である。

【図 3】 本発明の一実施の形態のプログラムの全体構成をブロック図形式で示す図である。

【図 4】 音声データの構成を模式的に示す図である。

【図 5】 図 3 に示す音響・韻律分析部 92 のブロック図である。

【図 6】 図 3 に示すケプストラム分析部 94 のブロック図である。

【図 7】 図 6 に示す規格化および統合部 144 のブロック図である。

【図 8】 図 3 に示すフォルマントの最適化部 98 のブロック図である。

【図 9】 図 3 に示す A Q 指数計算部 100 のブロック図である。

【図 10】 本発明の一実施の形態のプログラムによる表示例を示す図である。

【図 11】 音声データのうち、緊張音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

【図 12】 音声データのうち、氣息音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

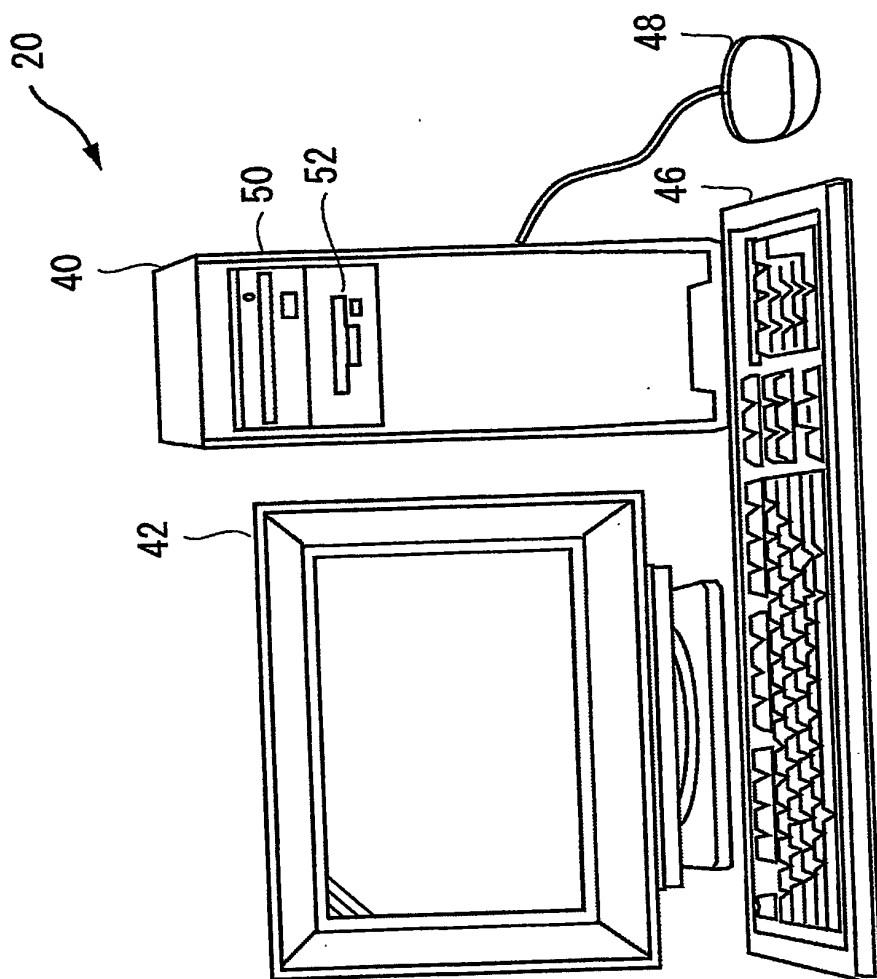
【図 13】 感知された氣息性と音響的に測定された A Q 指数との間の関連を示す散布図である。

# 【符号の説明】

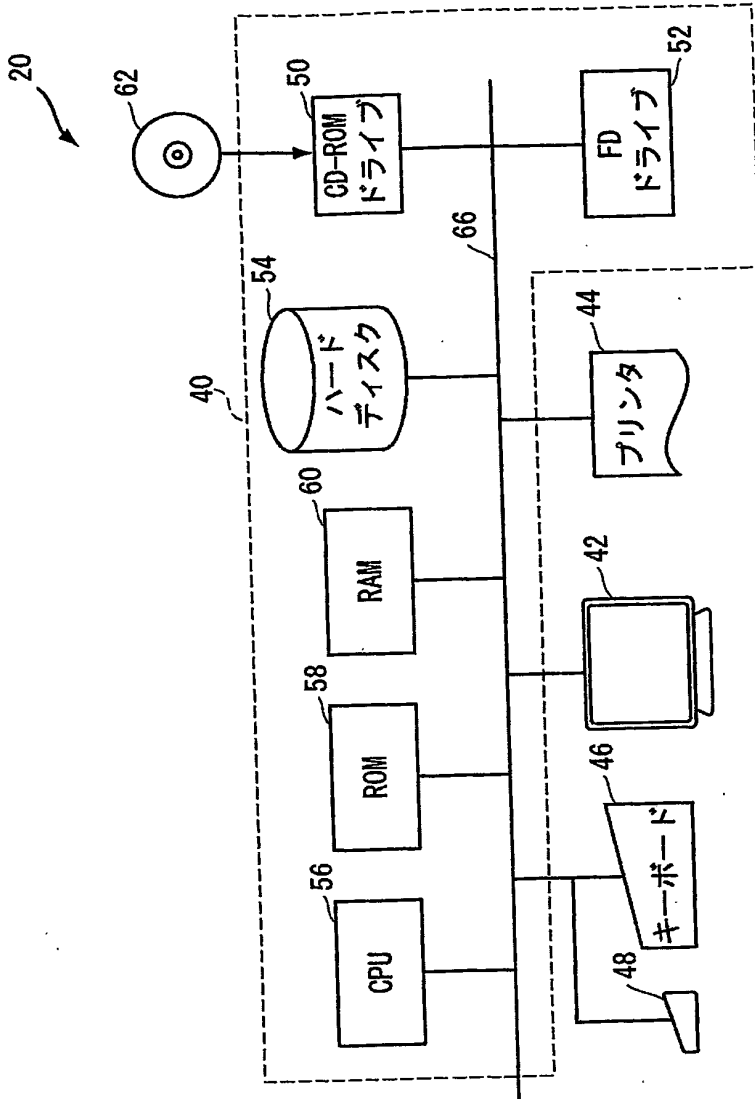
20 コンピュータシステム、82 音声データ、90 FFT 処理部、92 音響・韻律分析部、94 ケプストラム分析部、96 擬似音節中心の抽出部、98 フォルマントの最適化部、100 A Q 指数計算部、110 ピッチ判定部、112 ソノラントエネルギー算出部、114 ディップ検出部、116 有声・エネルギー判定部、130 線形予測分析部、132 フォルマント推定部、134  $\Delta$ ケプストラム算出部、136 ケプストラム再生部、138 フレーム間分散算出部、140 対数変換および逆 DCT 部、142 ケプストラム距離計算部、144 規格化および統合部、146 信頼性の中心候補出力部、186 距離最小化処理部

【書類名】 図面

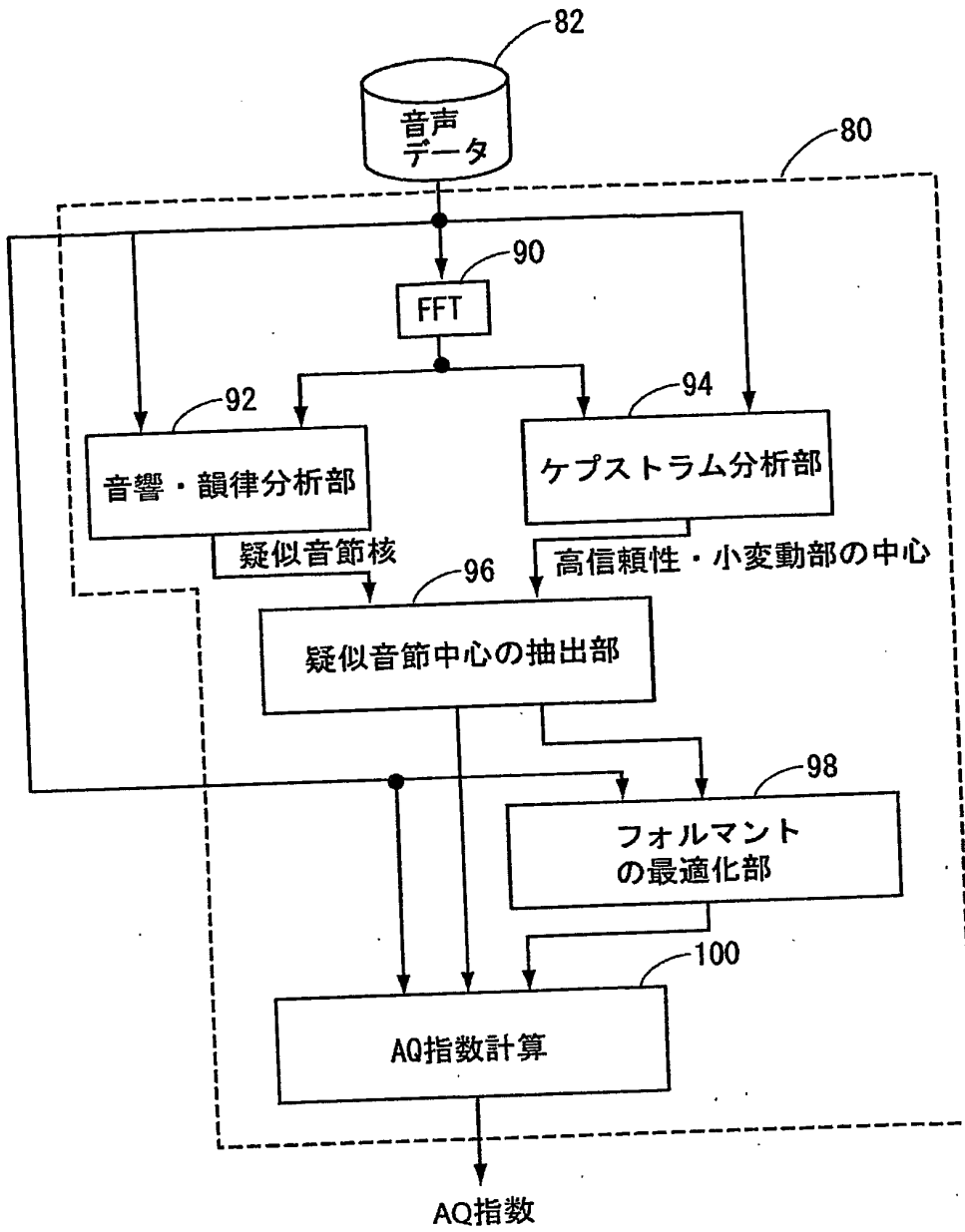
【図 1】



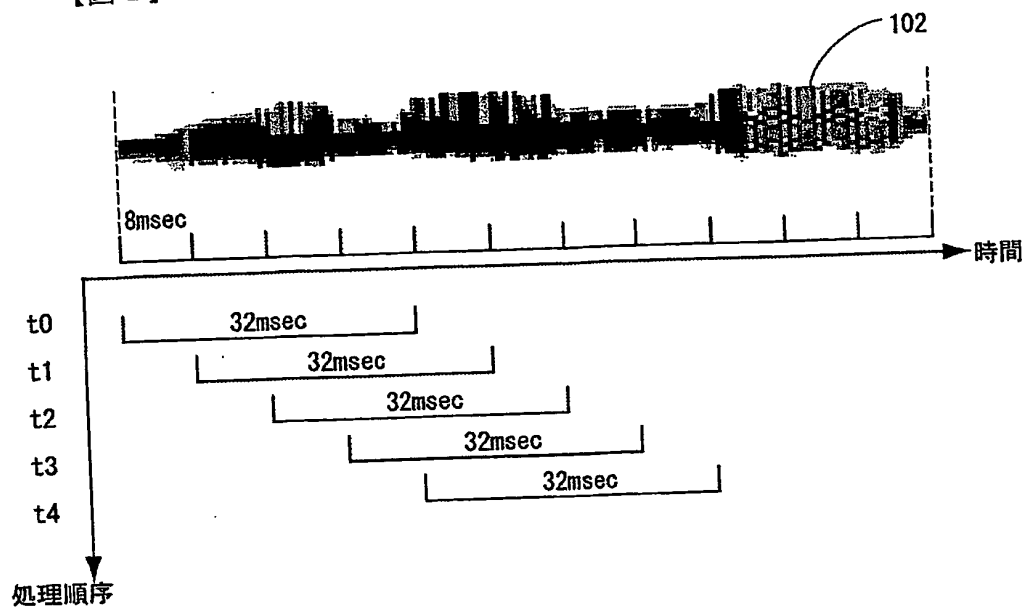
【図 2】



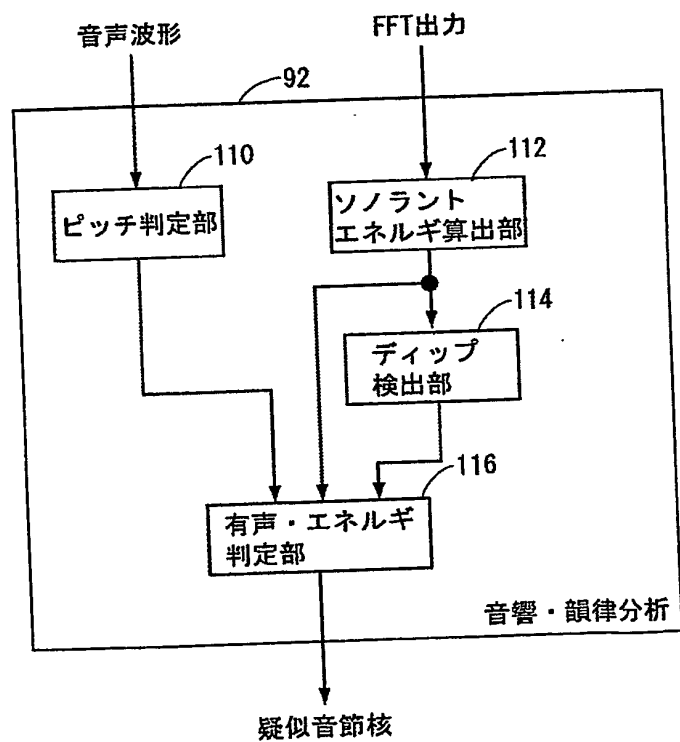
【図3】



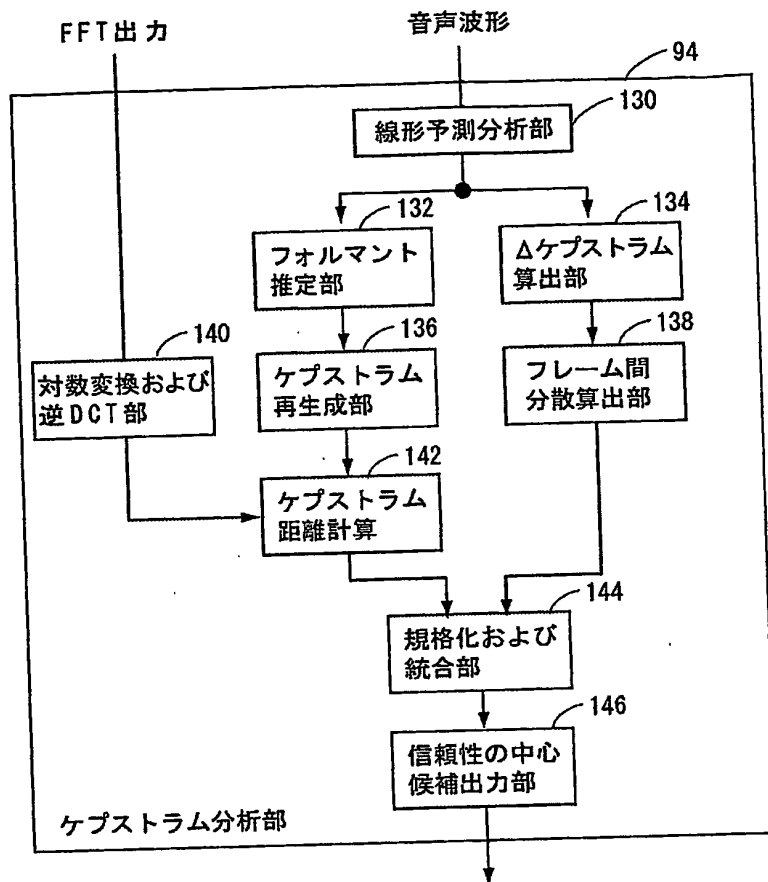
【図 4】



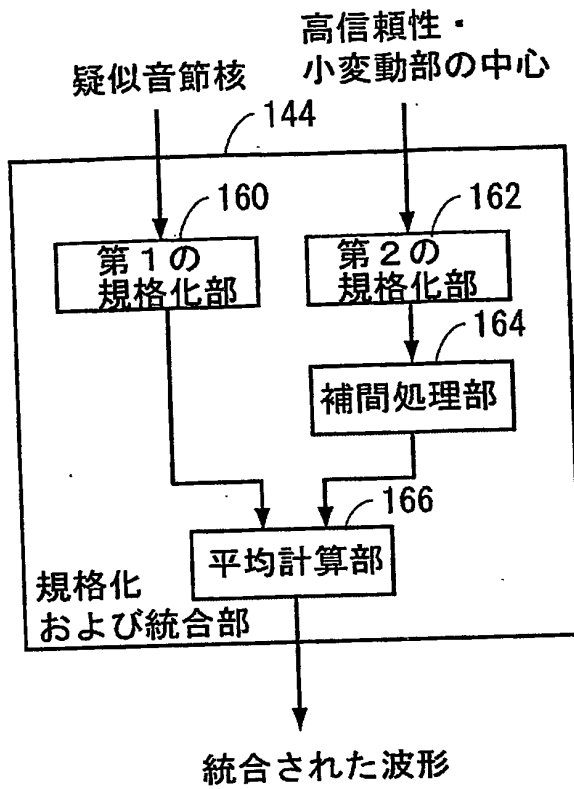
【図 5】



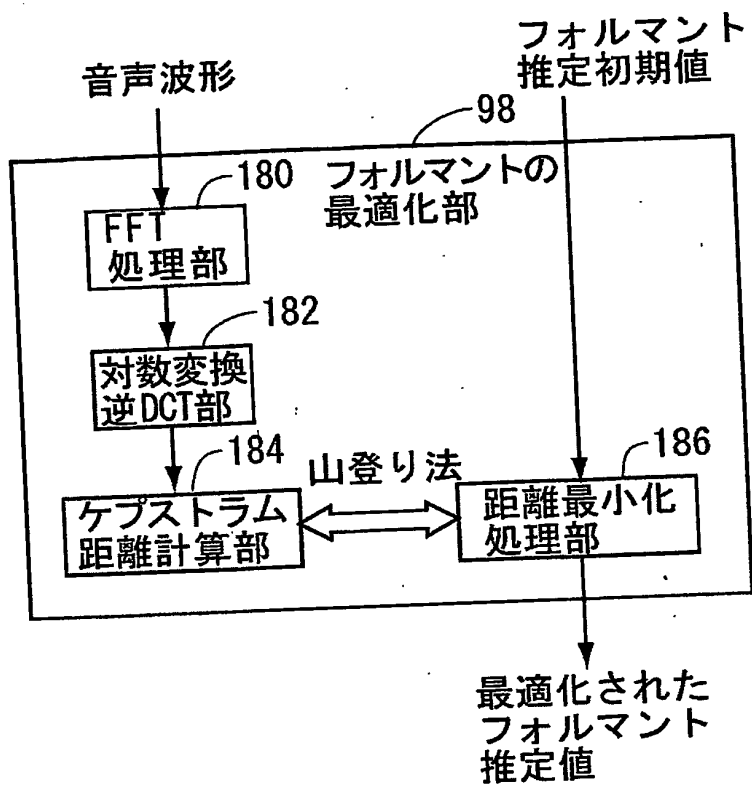
【図 6】



【図7】

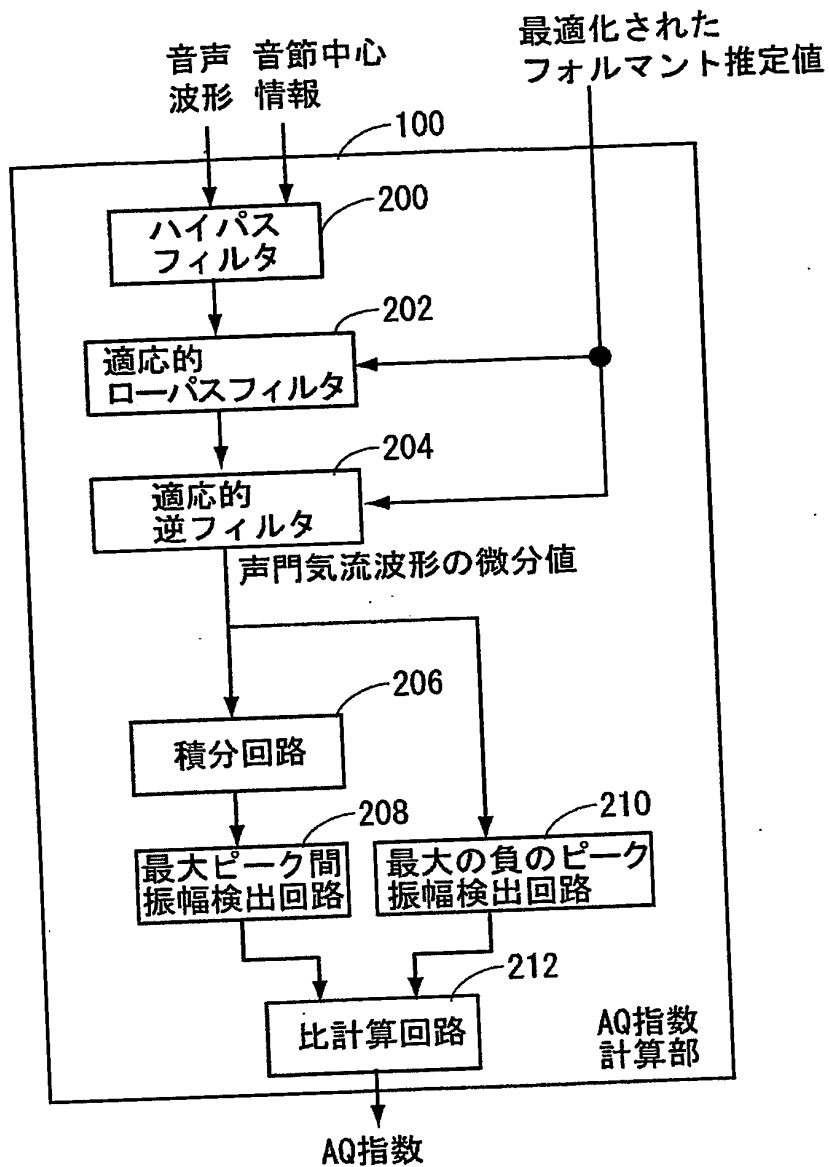


【図8】

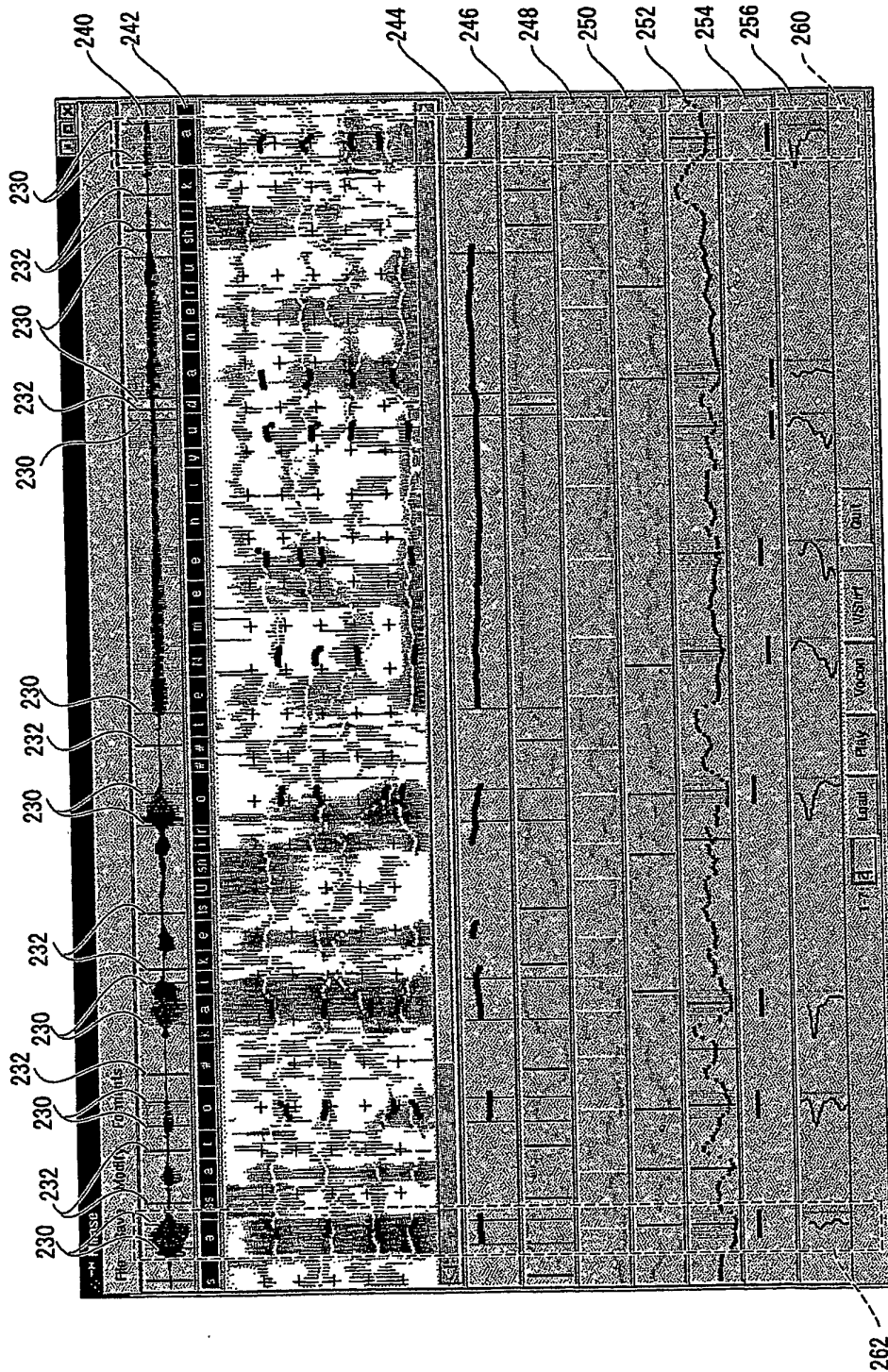




【図9】

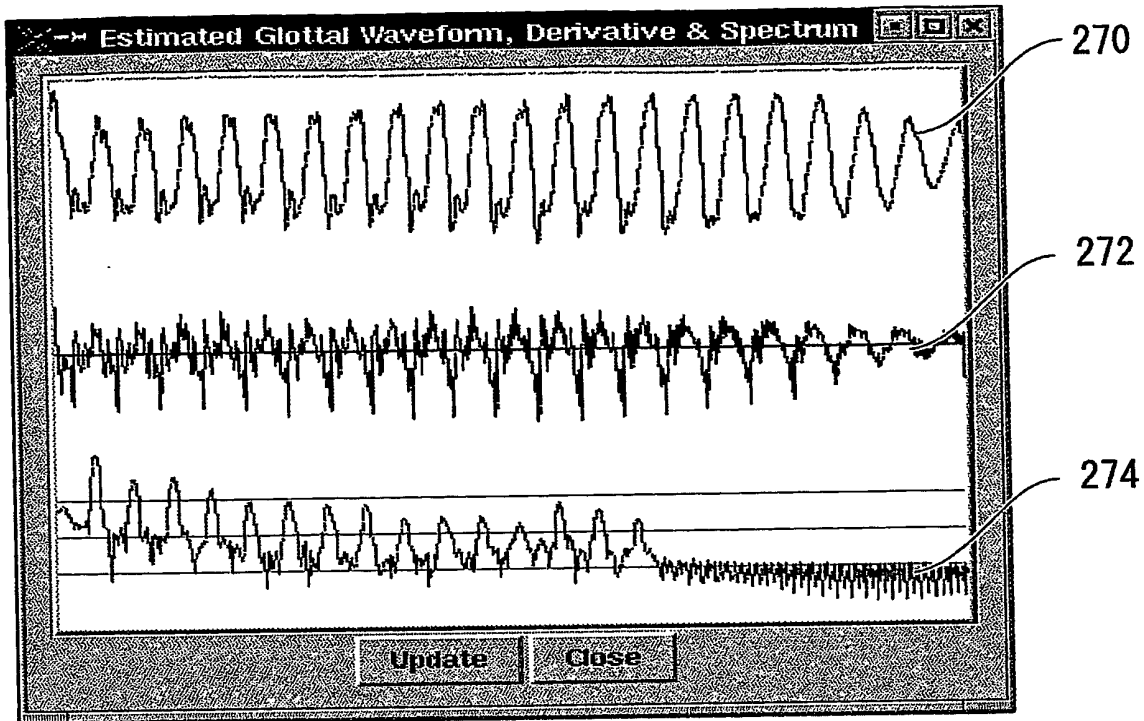


【図 10】

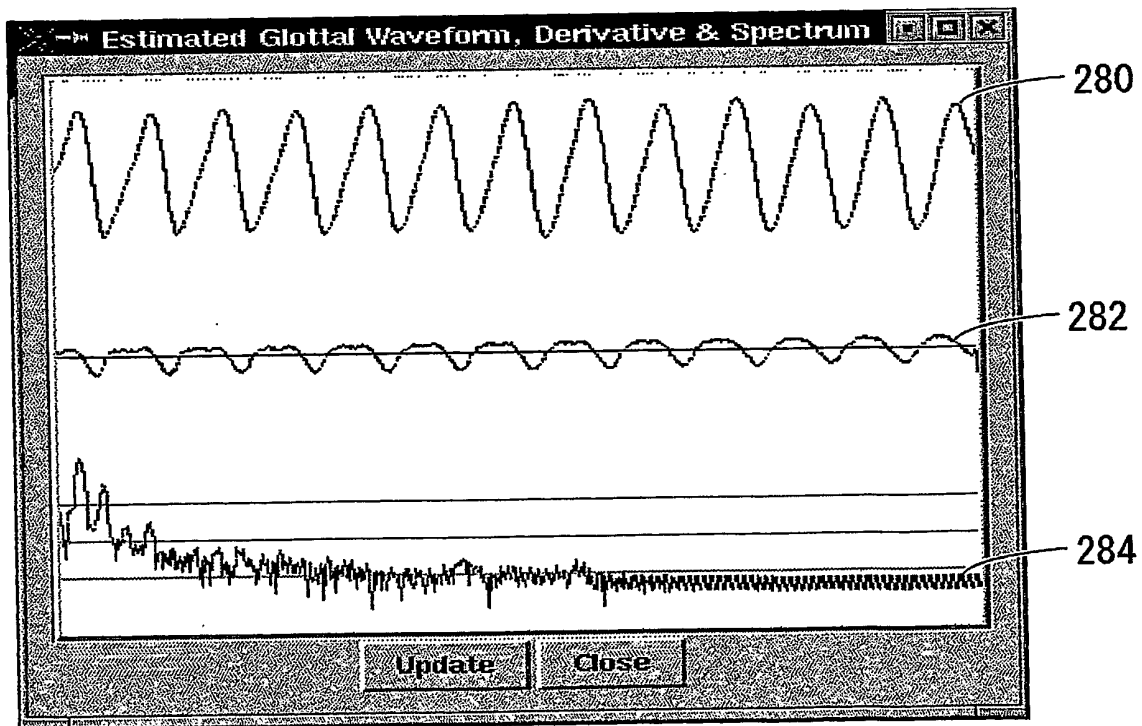


BEST AVAILABLE COPY

【図 1 1】

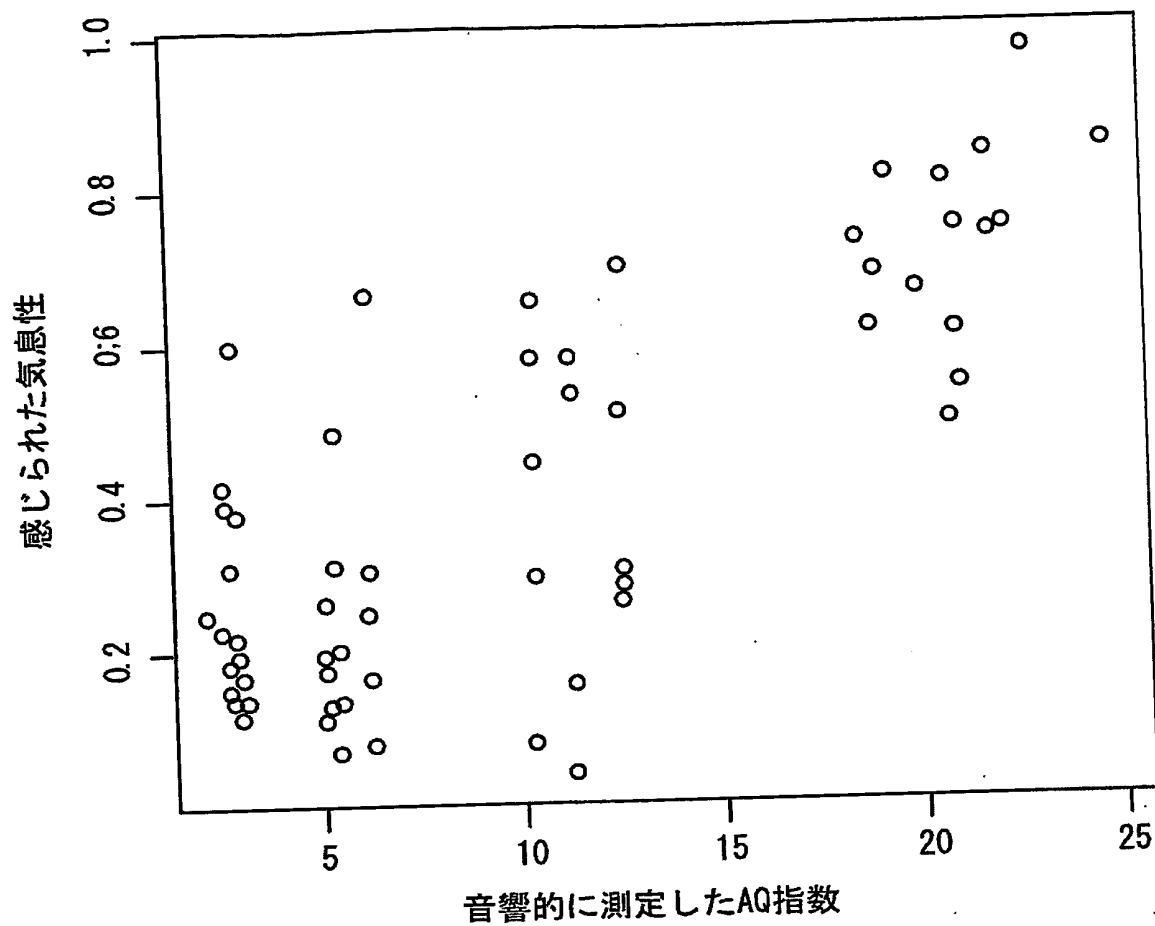


【図 1 2】



BEST AVAILABLE COPY

【図13】



【書類名】 要約書

【要約】

【課題】 音声波形の特徴を高い信頼性で示す部分を決定できるようにする。

【解決手段】 この装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち、話者によって安定して発生されている領域を抽出する音響・韻律分析部 9 2 と、データから音声波形のスペクトルの時間軸上の分布を算出し、その時間軸上の分布に基づいて、音声波形のうち、その変化が話者により良好に制御されている領域を推定するケプストラム分析部 9 4 と、話者によって安定して発生されている領域として抽出され、かつ話者によってその変化が良好に制御されていると推定された領域を音声波形の高信頼性部分として決定する擬似音節中心の抽出部 9 6 とを含む。

【選択図】 図 3

【書類名】 手続補正書  
 【整理番号】 0020001  
 【提出日】 平成15年 1月 9日  
 【あて先】 特許庁長官殿  
 【事件の表示】

【出願番号】 特願2002-141390

【補正をする者】

【識別番号】 396020800

【氏名又は名称】 科学技術振興事業団

【補正をする者】

【識別番号】 393031586

【氏名又は名称】 株式会社国際電気通信基礎技術研究所

【代理人】

【識別番号】 100099933

【弁理士】

【氏名又は名称】 清水 敏

【手続補正 1】

【補正対象書類名】 特許願

【補正対象項目名】 発明者

【補正方法】 変更

【補正の内容】

【発明者】

【住所又は居所】 埼玉県川口市本町4丁目1番8号 科学技術振興事業団  
 内

【氏名】 モクタリ パーハム

【発明者】

【住所又は居所】 京都府相楽郡精華町光台二丁目2番地2 株式会社国際  
 電気通信基礎技術研究所内

【氏名】 キャンベル ニック

【その他】 出願時発明者名の誤記を訂正するため

【プルーフの要否】 要

認定・付加情報

特許出願の番号	特願2002-141390
受付番号	50300025794
書類名	手続補正書
担当官	塩野 実 2151
作成日	平成15年 1月16日

<認定情報・付加情報>

【補正をする者】

【識別番号】

396020800

【住所又は居所】

埼玉県川口市本町4丁目1番8号

【氏名又は名称】

科学技術振興事業団

【補正をする者】

【識別番号】

393031586

【住所又は居所】

京都府相楽郡精華町光台二丁目2番地2

【氏名又は名称】

株式会社国際電気通信基礎技術研究所

【代理人】

申請人

【識別番号】

100099933

【住所又は居所】

大阪府大阪市北区西天満2丁目3番9号 オーク

西天満ビル3階 清水 敏特許事務所

【氏名又は名称】

清水 敏



出 願 人 履 歴 情 報

識別番号

[396020800]

1. 変更年月日	1998年 2月24日
[変更理由]	名称変更
住 所	埼玉県川口市本町4丁目1番8号
氏 名	科学技術振興事業団

出 願 人 履 歴 情 報

識別番号

[39303158.6]

- |          |                    |
|----------|--------------------|
| 1. 変更年月日 | 2000年 3月27日        |
| [変更理由]   | 住所変更               |
| 住 所      | 京都府相楽郡精華町光台二丁目2番地2 |
| 氏 名      | 株式会社国際電気通信基礎技術研究所  |